

Evaluating Large Language Models for Agricultural Injury Surveillance

Serap Gorucu¹, Jacob Muller¹, Daniel Petti¹, Changying Li¹, Matthew Pilz², Bryan P. Weichelt²

¹Department of Agricultural and Biological Engineering, University of Florida, Gainesville, Florida, United States

²National Children's Center for Rural and Agricultural Health and Safety, National Farm Medicine Center, Marshfield Clinic Research Institute, Marshfield, Wisconsin, United States

Abstract

Collecting and disseminating data for agricultural injury surveillance typically depends on manual input and human review, making the process both time-consuming and labor-intensive. A prime example is AgInjuryNews (AIN), a public platform that compiles injury reports from news articles and investigations. Because the content is unstructured, AIN currently relies on human reviewers to extract relevant information. However, the rise of Large Language Models (LLMs) offers a promising avenue for automation. This study explored the potential of LLMs to assist in the reviewer role at AIN. Models evaluated include OpenAI's ChatGPT 3.5 and 4, along with a fine-tuned version of Llama 2, to assess their accuracy in extracting incident and victim-related details. Each model was tasked with identifying specific data points such as drug or alcohol involvement, time of incident, and victim demographics from a random sample of news articles previously reviewed by AIN staff. The fine-tuned Llama 2 emerged as the top performer, with an average accuracy of 93% and perfect scores in some categories. While none of the models were perfect, the results highlight the feasibility of integrating LLMs to streamline workflows, reduce resource demands, and enhance the efficiency of data collection and analysis.

Key Words: *Injury Surveillance, Large Language Models, Language Processing, Automation*